

Species distribution model performance improves when habitat characterizations are centered on detected individuals instead of observers

Fang-Yu Shen^{a,*}, Fiona Victoria Stanley Jothiraj^b, Rebecca A. Hutchinson^{a,b},
Tyler A. Hallman^{a,c}, Jenna R. Curtis^{a,d}, W. Douglas Robinson^a

^a Department of Fisheries, Wildlife and Conservation Sciences, Oregon State University, Corvallis, OR 97331, USA

^b School of Electrical Engineering and Computer Science, Oregon State University, Corvallis, OR 97331, USA

^c School of Environmental and Natural Sciences, Bangor University, Bangor LL57 2DG, UK

^d Cornell Lab of Ornithology, Cornell University, Ithaca, NY 14850, USA

ARTICLE INFO

Keywords:

Specialist
Model performance
Abundance
Positional error
Landsat
Effective detection radius
Habitat characterization

ABSTRACT

Species distribution models (SDMs) link species occurrence to environmental characteristics to predict suitable habitats beyond known occurrences. The conventional procedure to fit SDMs for individual organisms detected at some distance away from observers is to characterize species' associated habitat based on observer's survey location. However, each surveyed individual may be detected in habitats distinct from those where observers are located. Here, we compared environmental variables centered on the observer and individual bird locations and the consequent effects on SDMs performance. We utilized remote sensing data on observer- and bird-locations to characterize habitat at three radii (pixel radius: 30-m; fixed radius: 100-m; species-specific effective detection radius). We trained Poisson boosted regression tree models for 105 bird species from structured professional surveys. We evaluated models' predictability with Kendall's rank correlation coefficient and used linear mixed-effect models to measure the effect of characterization locations and radii. Models based on bird locations exhibited a median increase of 22.9% in predictive performance, demonstrating higher Kendall's rank correlation coefficients than those based on observer locations, leading to more reliable prediction maps. SDMs of habitat specialists and generalists performed better when habitat characterization was centered on bird instead of surveyor locations. A higher percentage of habitat specialists (72%) than generalists (55%) showed better model performance in bird-location than in observer-location models. Across radii, fixed radius generally performed better than species-specific effective and pixel radii. Our findings emphasize the importance of prioritizing habitat characterizations based on detected individuals' locations to enhance model performance and improve species distribution predictions.

1. Introduction

The distributions, population sizes and functioning of biodiversity have been influenced by anthropogenic factors from local to global scales (Matuoka et al., 2020; Betts et al., 2022; Williams and Newbold, 2020; Gaston et al., 2003). Establishing reliable measurements of species distributions is essential to conservation and management to mitigate such changes (Whytock et al., 2018; Olea and Mateo-Tomás, 2011; Gábor et al., 2024; Robinson et al., 2021). Species distribution models (SDMs) are commonly used to interpolate or predict species occurrence, and sometimes abundance, information by analyzing environmental (habitat and climate) data associated with species detections and non-

detections (Elith et al., 2008; Barbet-Massin and Jetz, 2014; Adde et al., 2023).

Sample size, sampling bias, and species' specialization (i.e., generalists, specialists) are common factors known to influence the performance of SDMs (Hallman and Robinson, 2020; Moudrý et al., 2024). However, positional error is rarely addressed: in most wildlife surveys, individuals are not detected at observers' exact location (Moudrý et al., 2024; Zhang et al., 2018). Multiple factors contribute to positional error, including recording techniques, spatial resolution, and the characteristics of the species being surveyed (Moudrý et al., 2024). Most data supplied to SDMs utilize habitat data centered on where a survey was conducted (i.e., observer-location) (Hallman and Robinson, 2020;

* Corresponding author.

E-mail address: bettysheen@gmail.com (F.-Y. Shen).

Shirley et al., 2013; Betts et al., 2014). Despite some studies showing a positive or little impact of small positional errors on fitting SDMs (Smith et al., 2023; Fernandez et al., 2009), several studies have shown a negative impact of positional error on SDMs performance (Gábor et al., 2023; Gábor et al., 2022; Gábor et al., 2020; Osborne and Leitão, 2009). Birds, for example, are typically detected at some distance from an observer and may, therefore, be in habitats different from those immediately around the surveyor. This is particularly true in surveys of heterogeneous landscapes where observations are gathered along roads passing through patchy arrangements of habitats (Moudré et al., 2024). Thus, characterizing environmental predictors around observer locations to construct SDMs can induce mismatches between habitats used by birds and those directly around the observer (Guisan and Thuiller, 2005). Possible mismatching of observer's habitat with bird's habitat will produce incorrect model output and can reduce model performance (Osborne and Leitão, 2009), perhaps more so for habitat specialists than generalists (Gábor et al., 2020). The importance of improving model performance could be large because many structured avian surveys have been conducted along major roadways, such as the North American Breeding Bird Survey (Ziólkowski et al., 2022), and provide the foundation for regional and continental estimates of population sizes (Edwards et al., 2023). Few studies have quantified this issue of locational mismatch despite widespread use of point count protocols where observers estimate distance and direction of birds from the observer located at a survey point.

In addition to researching the influences of positional error when fitting SDMs, characterizing habitat elements and utilizing relevant spatial scales to predict species distributions are key steps in fitting SDMs (Hallman and Robinson, 2020; Moraga et al., 2019). Environmental predictors can originate from various spatial scales, typically encompassing larger scales for climatic predictors and smaller scales for habitat configuration. Many studies have employed a combination of large and small scales of environmental predictors to fit SDMs (Goetz et al., 2014; Burns et al., 2020; Jan et al., 2025). Ideally, the resolution of environmental predictor variables should be at scales most influential to species' presence or abundance. Moudré et al. (2023) reviewed the literature on selecting different spatial resolutions for predictor and response variables. Multiple studies indicate that model performance declines with the coarsening spatial scale of predictor variables. Regarding spatial scale selection, species usually respond to environment differently at various spatial scales, and it is common to find that no single scale applies best to all species (Connor et al., 2018; Mitchell et al., 2001). Suitable spatial radii may be large (typically < 2000 m) for mobile species and those with large home ranges such as birds. The optimal radius to infer bird-habitat relationships may vary among bird species. Gábor et al. (2024) tested four spatial grain sizes (1 km² to 2500 km²) on 57 water-associated bird species and found that 1 km² grain size yielded the best model performance. Hallman and Robinson (2020) compared single-scale and multi-scale approaches from 165 m to 1215 m radii and found that multi-scale performed significantly better than single-scale in 25 bird species.

Estimating precise coordinates of birds from observers is a challenging enterprise, largely because many birds are heard and not necessarily seen. Weather, habitat structure, time of day, characteristics of vocalizations and even distance itself affect detectability of birds (Edwards et al., 2023; Amundson et al., 2014; Farnsworth et al., 2002; Shen et al., 2023; Marsh and Sinclair, 1989; Anderson et al., 2015; Diefenbach et al., 2003). Generally, birds are less detectable as distance from an observer increases (Buckland et al., 2007; Reynolds et al., 1980). Taking advantage of this ubiquitous pattern of change in detectability as a function of distance, we can estimate density and other related factors such as absolute abundance or population sizes (Farnsworth et al., 2002; Buckland et al., 2007). The standard approach for such analyses is to generate a detection function $g(r)$, which characterizes the sigmoidal relationship in numbers of individuals (typically dozens to hundreds of observations) detected as a function of estimated

or measured distance from the observer (Buckland et al., 2015). From analysis of the detection function, the effective radius (μ) can be calculated, which is the distance at which the estimated number of missed individuals beyond that distance equals the number of detected individuals less than that distance (Anderson et al., 2015; Buckland et al., 2015). The effective radius is species-specific and affected by measured distance between observers and birds. Effective radii have been recently applied to data from autonomous recording units (ARUs), for example, to aid in estimation of density (Pankratz et al., 2017; Pérez-Granados and Traba, 2021). Consequently, species-specific effective radii used in modeling species distributions may better inform characterization of the habitats and other environmental predictors that species occupy.

We calculated species-specific effective radii for breeding bird species in Oregon, U.S.A., from more than 10,000 surveys with direction and distance data. Positional error is most likely to be relevant when the resolution of predictors is significantly affected by it. Thus, we compared the predictive performance of SDMs built with habitat variables centered on estimated location of each bird to those centered on observer locations. For each set of habitat variables, we analyzed three radii: (1) pixel size – 30 m spatial resolution of Landsat; (2) fixed radius – 100 m, a commonly used radius in bird surveys; and (3) species-specific effective radius, which recognizes that effective detection of each bird varies as a function of species-specific traits such as loudness of vocalizations or conspicuousness of behaviors. We also measured the effects of positional error and three radii sizes on SDMs performance. We anticipate that positional error will be most influential at the smallest scale and less so at the 100-m radius scale. Finally, we tested whether habitat specialist species obtain better SDMs predictions than habitat generalist species when characterizing habitat variables on bird instead of surveyor locations.

2. Methods

2.1. Study area

Our study area was the state of Oregon, U.S.A. Oregon possesses an area of 255,026 km², and has high habitat diversity and climatic variation. For example, it includes 12 Köppen climate types and 9 ecoregions (Hopkins et al., 2022). Temperature variability has ranged from -48 to + 50 °C in the past 150 years. Habitats include deserts to rain forests, influenced by annual rainfall varying from 0.1 to 5 m/year across its geography. Elevations extend from sea level at the Pacific Ocean to 3427 m at its highest peak (Mt. Hood). In general, major habitat types include sagebrush shrubland, forests, grasslands, agriculture, and suburban and urban land (Robinson et al., 2020).

2.2. Bird data

Throughout this study, we used two datasets derived from the Oregon 2020 project, both were gathered by the same group of trained observers. The first dataset included the direction and distance to each bird. The second dataset was derived from the first, but with the distance information removed. Since the second dataset was gathered in the same way data submitted to the public archive – eBird, we refer to these stationary counts following the eBird stationary count protocol as “eBird-Oregon 2020.” Only the first dataset of Oregon 2020 data contained distance measurements of distance between the observer and each bird, so we applied distance sampling algorithms to estimate species-specific effective radii for 198 breeding bird species and used the distance data along with direction from an observer to estimate the coordinates of each bird's location. When comparing SDMs' performance for bird- vs. observer-location, we included the first dataset of Oregon 2020 (distance information included) for bird-location model training. In contrast, eBird-Oregon 2020 (lacking distance information) was used for observer-location model training. Utilizing these two

subsets of data in this way, we avoided issues of radii around bird- and observer-locations overlapping so much that detection of habitat differences was impeded. For further SDM performance testing, we classified species into habitat specialists versus generalists based on breeding evidence during surveys and species natural history information from Oregon (Marshall et al., 2003). We used bird occurrence records from the Oregon 2020 project (Robinson et al., 2020), and selected 198 breeding bird species (Table S1) with sufficient numbers of detections for further modelling procedures. Oregon breeding bird species were defined as species that breed in the region, including year-round residents, and migratory species that breed in Oregon during spring and summer (Marshall et al., 2003). Oregon 2020 sampling points were distributed across Oregon in a stratified random manner (Fig. 1). A total of 10,844 bird surveys were conducted from 15 May to 10 July 2011–2019, with earlier counts conducted in sites of lower elevation, and later counts conducted at higher elevation sites. Surveys were conducted from dawn to noon by trained observers (Farmer et al., 2012). During each survey, the observer first used *Garmin eTrex 10* to record their location ($\pm 15\text{m}$) and recorded all birds seen and heard during 5-min stationary counts in a radius of unlimited distance. The distance of all individuals from the observer were measured with a range finder to the nearest 10 m when the birds were seen. We also measured distance to the tree or other structure a bird was considered to be present in ($\pm 10\text{m}$) even when we could not see the bird, a standard practice in avian point counting. Direction was indicated as one of 16 different possibilities (e.g., north, north-northeast, northeast, east-northeast, etc.). Using distance from an observer and the direction of the bird, we deployed *Geopy*, a Python library used to perform Vincenty's Formulae (Vincenty, 1975), treating Earth as an ellipsoidal geometry to estimate coordinates (i.e., latitude and longitude) where each bird occurred. Thus, Oregon 2020 data allowed us to build species-specific distribution models based on both observer locations and estimated bird locations. To fit SDMs with detection vs. non-detection data records, we filled in non-detected species with zero counts (i.e., zero-filled) in each survey based on 198 bird species included in this study.

2.3. Environmental data

Recent studies have implemented Landsat spectral bands in fitting SDMs and predicting bird distributions (Shirley et al., 2013; Hopkins et al., 2022; Jefferys et al., 2024). To characterize habitats, we retrieved six Landsat 7 Surface Reflectance Collection spectral bands – Bands 1–5 & 7 (Table 1), from May 1 to July 31, 2011–2019. Cloud cover may impact the spectral bands collection in remote sensing (Zurell et al., 2020). To reduce the impact of clouds on reflectance values, we

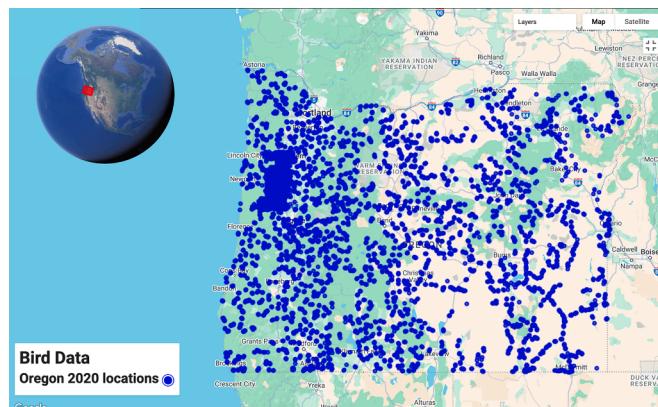


Fig. 1. Distribution of survey sites from Oregon 2020 (10,844 sites) in the state of Oregon, USA. Oregon 2020 data was used for bird-location SDMs training; whereas eBird-Oregon 2020 (subset of Oregon 2020 lacking distance information) was used for observer-location SDMs training.

Table 1

Landsat 7 Enhanced Thematic Mapper Plus (ETM+) spectral bands description and ecological inferences (Landsat 7 Data Users Handbook; What are the best Landsat spectral bands for use in my research).

Band Name (abv.)	Scale (m)	Description (wavelength)	Ecological Inferences
Blue Band (SR_B1)	30-m	Blue light surface reflectance (0.45–0.52 μm)	Differentiating rock types from vegetation and distinguishing deciduous from coniferous vegetation. This band also helps identify areas of shallow water, where living plants reflect deeper blue-violet hues.
Green Band (SR_B2)	30-m	Green light surface reflectance (0.52–0.60 μm)	Assessing vegetation health, this band is sensitive to chlorophyll gain or loss. Higher reflectivity of the Green Band indicates chlorophyll loss, because leaves absorb less energy. The SR_B4/SR_B2 ratio is used to detect senescing vegetation.
Red Band (SR_B3)	30-m	Red light surface reflectance (0.63–0.69 μm)	Discriminating vegetation types. This band helps to assess the content of chlorophyll of vegetation. Higher chlorophyll in vegetation absorbs energy, thus the reflectance of Red Band is low. The SR_B4/SR_B3 ratio is used to detect growing vegetation.
Near-infrared Band (SR_B4)	30-m	Near-infrared surface reflectance (0.77–0.90 μm)	Emphasis on biomass content. High SR_B4 reflectance indicates healthy vegetation, because water in leaves reflects light.
Shortwave Infrared 1 (SR_B5)	30-m	Shortwave infrared 1 surface reflectance (1.55–1.75 μm)	Distinguishing rocks and soils from vegetation and soil, measuring moisture content. The SR_B4/SR_B5 is used to detect soil content.
Shortwave Infrared 2 (SR_B7)	30-m	Shortwave infrared 2 surface reflectance (2.08–2.35 μm)	In addition to detecting moisture content of soil and vegetation, it measures rocks and mineral formation altered by hydrothermal processes.

performed a cloud masking technique from "QA_PIXEL" band to remove cloud coverage. Since bird surveys occurred from May into July, we composited all Landsat imagery from May to July to generate a median value for each annual image composite across Oregon (Jefferys et al., 2024). We measured habitat characteristics as close to the date of each observation as possible (Araújo et al., 2019; Crego et al., 2022). For example, an observation made in 2012 was matched with the Landsat annual composite from 2012. All environmental analysis was performed in Google Earth Engine (<https://earthengine.google.com/>), a cloud-based application programming interface for geographic information system analysis, which has recently been applied to predict species distributions (Crego et al., 2022).

Each observation from the Oregon 2020 data was referenced to two habitat characterizations (observer-location & bird-location). For each location type, we also summarized habitat by calculating the median of each Landsat band at three radii: a) using exact coordinates to retrieve six Landsat bands at a buffer of 30-m pixel radii; b) creating a fixed buffer of 100-m around each location; and c) applying a species-specific effective radius buffer around each coordinate (Fig. 2).

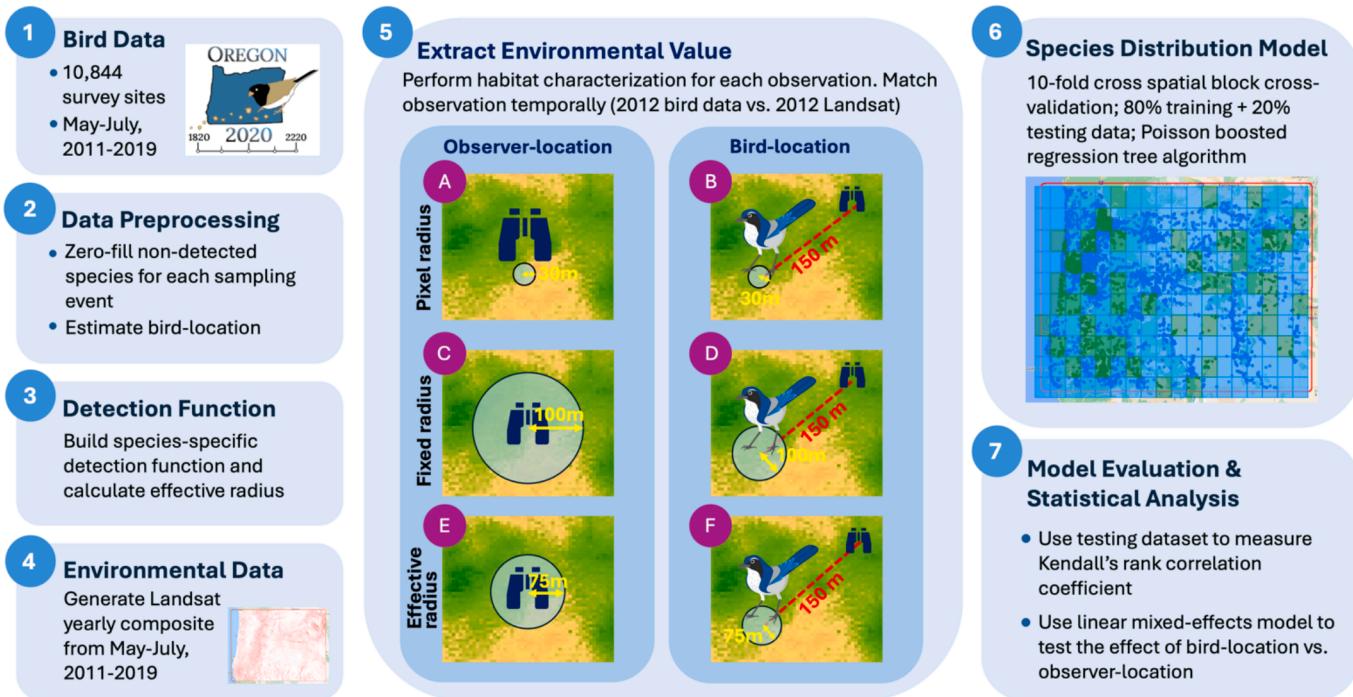


Fig. 2. Schematic of data processing to compare different habitat characterization approaches on fitting species distribution models. Numbers in each image's top-left corner indicate the data processing order.

2.4. Statistical analysis

2.4.1. Effective radius

To assess the effect of distance on perceptibility, we employed distance analysis through the *Distance* package in R, building species-specific detection functions via distance information obtained from the Oregon 2020 project. Previous study of the relationships between using default half-normal versus species-specific key functions to fit detection curves has shown that using species-specific key functions yields better model fits (Clements et al., 2025). Therefore, to assess fit of each model, we employed three key functions fitting the distance data – a) hazard-rate; b) uniform; and c) half-normal with adjustments terms (i.e., Cosine, Hermite, and simple polynomial) for each bird species. We selected the most parsimonious model via Akaike's Information Criterion (AIC) and estimated average perceptibility of each species. Based on the best-selected model, we calculated effective radius for each species as follows:

$$\mu = \omega \times \sqrt{P_a} \quad (1)$$

where ω is truncation distance (m), and P_a is estimated perceptibility of species a . To avoid heterogeneity of detection probability that may affect fitting detection functions, we set truncation distance at 10 % less than the largest detected distance of a species, as suggested by (Buckland et al., 2015; Edwards et al., 2023).

2.4.2. Species distribution model

Six boosted regression tree (BRT) models were built for each species, associated with habitat characterization strategies on observer-location and bird-location (Fig. 2). To assess the association between bird abundance and Landsat spectral bands, we built Poisson BRT models for each species using the *dismo* package in R. The modelling process is as follows. All detection and non-detection observations of each species were split so that 80 % of the data was used for training and 20 % for evaluating the models. Thus, species' individual counts were treated as response variables and data from the six Landsat spectral bands were treated as explanatory variables. One geographic region, Benton and

Polk Counties, had especially dense sampling. To reduce the effects of oversampling in that region, for each observation we used the "spatio-temp_weight" from the *dynamicSDM* R package and set the radius within which we adjusted relative sampling effort to 104200 m, which was the longest distance between the center of that region and its most outlying sampling site.

At the scale of the entire state, we used "spatiotemp_block" from *dynamicSDM* R package to spatially split the training dataset with a $1^\circ \times 1^\circ$ block of Oregon based on the correlation of all Landsat spectral bands (Valavi et al., 2019; Bagchi et al., 2013). Each block was assigned a fold number from 1 of 10 to set up 10-fold cross-validation. The spatial block approach allows the training dataset to be spatially separated from the testing dataset (American Robin example shown in Fig. S1). Each spatial block possessed approximately equal sample size of survey points. A total of 5000 iterations were performed before assigning each observation to a final spatial block. We left out each block in turn as a test dataset and fitted BRT models to the remaining nine blocks of training data. For tuning parameters in all BRT models, we held a constant learning rate of 0.001, a tree complexity of 5, and an optimal number of trees automated from "gbm.step".

To calculate model performance on the evaluation datasets for each habitat characterization strategy of all bird species (i.e., predicted vs. observed values), we used Kendall's rank correlation coefficient (Hallman and Robinson, 2020). To better understand the differences in model performance between bird- and observer-location SDMs, we calculated the percentage difference for each species (Hallman and Robinson, 2020). The percentage difference was calculated by subtracting observer-location from bird-location evaluation metrics, and dividing by the evaluation metric of the observer-location model for each species. By focusing on the percentage difference relative to the observer-location model's evaluation metric, this approach reduces species-specific effects and facilitates a more standardized comparison across species for our three habitat characterization radii. Finally, we used linear mixed-effects models to evaluate the effects of characterization locations (i.e., bird-location models vs. observer-location models) and radius of environmental predictors on performance as evaluated by Kendall's rank correlation. We included an interaction

effect between characterization locations and radii to allow the model to determine more flexible effects and included species as a random effect. We excluded the results of SDMs that did not successfully complete the training process (i.e., failed model fitting). To visualize the predicted results from different habitat characterization strategies, we chose observer-location pixel, fixed radii (commonly used methods in SDMs), and bird-location fixed-radius BRT models of five species from habitat generalists and specialists, respectively, to predict on the 2019 Landsat spectral bands composite.

3. Results

We successfully fitted detection functions for all 198 breeding bird species. For fitting SDMs, a total of 105 of the original 198 bird species were successfully trained with observer- and bird-location models, whereas 93 species failed to fit SDMs in either observer- or bird-location models, likely due to insufficient numbers of observations (Table S1). Thus, we omitted species that failed to fit SDMs in subsequent analyses. We summarized results into three folds as follows: (1) effective radius, (2) SDMs of all species, and (3) SDMs of habitat generalists and specialists.

3.1. Effective radius

The median detected distance of individual birds across all our distance data was 80 m (Fig. S2). The most frequently selected key function of detection function via AIC was hazard-rate (54 species), followed by uniform (43 species) and half-normal (8 species). For species-specific effective radius, Long-billed Curlew (*Numenius americanus*) had the highest at 339 m (95 % CI = 271–407 m), followed by Pileated Woodpecker (*Dryocopus pileatus*) at 316 m (95 % CI = 269–362 m), and Common Nighthawk (*Chordeiles minor*) at 308 m (95 % CI = 254–363). In contrast, the three lowest effective radius species are Cliff Swallow (*Petrochelidon pyrrhonota*) at 26 m (95 % CI = 19–33 m), followed by Red-breasted Sapsucker (*Sphyrapicus ruber*) at 24 m (95 % CI = 13–35 m), and Mallard (*Anas platyrhynchos*) at 19 m (95 % CI = 3–36 m) (Fig. S3).

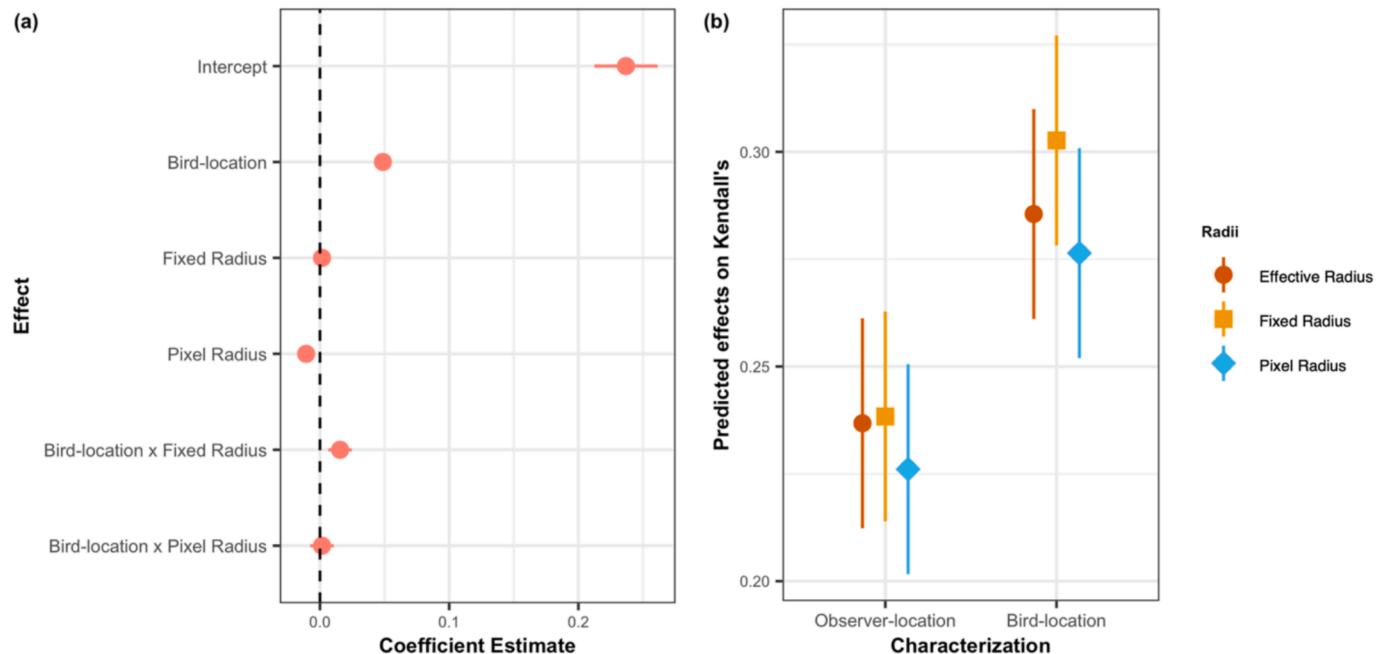


Fig. 3. Influence of habitat characterization locations and radii on Kendall's rank correlation coefficient from linear mixed-effect models. Panel (a) shows coefficient estimates of bird-location models on observer-location models from fixed and interaction effects. The effect of a variable on Kendall's is regarded as significant when the 95% CI did not overlap zero (vertical dashed line). Panel (b) shows the predicted effects of habitat characterization locations and radii on Kendall's rank correlation coefficient.

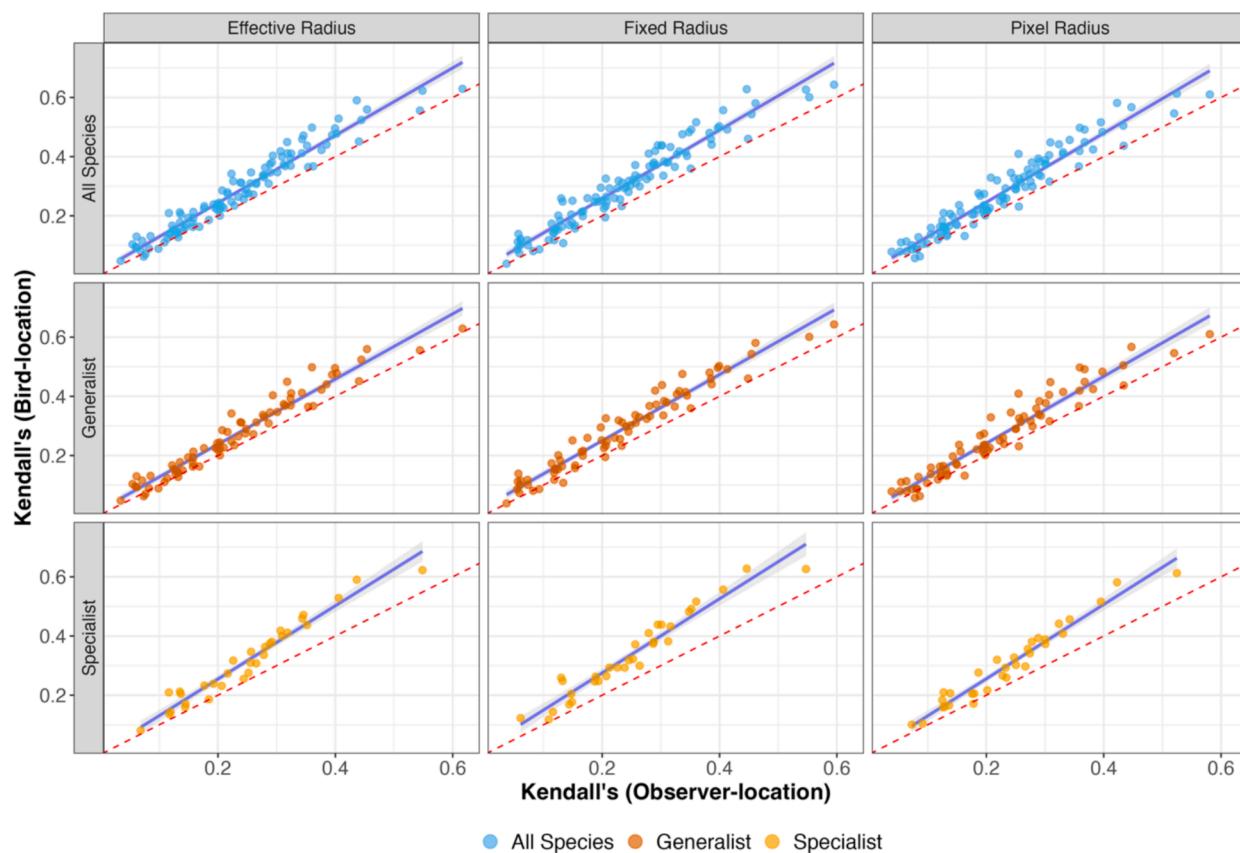


Fig. 4. Kendall's rank correlation coefficient of bird-location distribution models as a function of the observer-location distribution models based on species specialization and radii (1:1 plot). The columns are radii size for habitat characterization. Effective radius is a species-specific distance value; fixed radius is 100 m; and pixel radius is 30 m. The rows are species specialization categories – all species ($n = 105$), habitat generalists ($n = 73$), and habitat specialists ($n = 32$). Each point indicates a species. Red dashed line denotes the identity line where Kendall's values are equal from both bird- and observer-location models. Blue solid line with gray-shadowed 95 % CI represents the estimated slope of Kendall's values between bird- and observer-location models. Points that fall above the identity line indicate a higher model performance in bird-location than observer-location models. In contrast, points that fall below the identity line indicate a lower model performance in bird-location than observer-location models. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

showed that bird-location models outperformed observer-location models in 23/32 (72 %) of the specialists (Fig. S7). Habitat specialists using a fixed radius (100 m) to characterize habitat on bird-location produced the largest model percentage improvement compared to observer-location models (median = 36.7 %) (Fig. S4c). Across three radii, pixel radius models showed a negative effect on Kendall's rank correlation coefficient (Fig. S6a) and led to an overall lower model performance and percentage improvement (Fig. S4c & S5c). Utilizing different habitat characterization methods yielded different predicted relative abundance and evaluation metric values across specialists (Fig. 5). Overall, bird-location models demonstrated better model performance, indicating that the predicted values more closely align with the observed abundance of a species. Despite these differences in predicted relative abundance, most specialists exhibited similar predicted distribution patterns throughout the state across three habitat characterization strategies. However, the predicted maps for the Wrentit (*Chamaea fasciata*) displayed the greatest contrast between observer- and bird-location models in the western region (Fig. 5).

For habitat generalists, Kendall's rank correlation coefficient had a median of 0.26 across bird-location and a median of 0.21 for observer-location models (Fig. S5b). The results of linear mixed-effect model showed that bird-location models outperformed observer-location models in 40/73 (55 %) of the generalists (Fig. S8). For bird-location models, SDMs that applied fixed radius to characterize predictors produced the largest model percentage improvement compared to observer-location models (median = 24.11 %) (Fig. S4b). Similar to specialists,

pixel radius models showed a negative effect on Kendall's rank correlation coefficient (Fig. S6) and led to an overall lower model performance and percentage improvement (Fig. S4b & S5b). The prediction maps for generalists produced similar conclusions as specialists, where bird-location models showed better performance than observer-location models (Fig. 6).

4. Discussion

To compare model performance when predictors were centered on observer location versus location of surveyed individuals, in our case birds, we measured Kendall's rank correlation coefficient among boosted regression tree (BRT) models for 105 breeding bird species, including 73 habitat generalists and 32 habitat specialists. We characterized habitat via three radii – a) species-specific effective radius; b) fixed radius: 100 m; and c) pixel radius: 30 m, by matching species abundance with yearly Landsat imagery composites. Our findings in Kendall's rank correlation coefficients suggest that characterizing habitat based on locations of surveyed individuals leads to more accurate model outputs. A higher model accuracy indicates a more reliable distribution and abundance prediction, a useful outcome for conservation practitioners. These results are consistent with those of Gábor et al. (2023) who investigated the impact of positional error with virtual species and a real species – Band-tailed Pigeon (*Patagioenas fasciata*); their findings showed model performance generally decreased as positional error increased yet stated that models derived from positionally

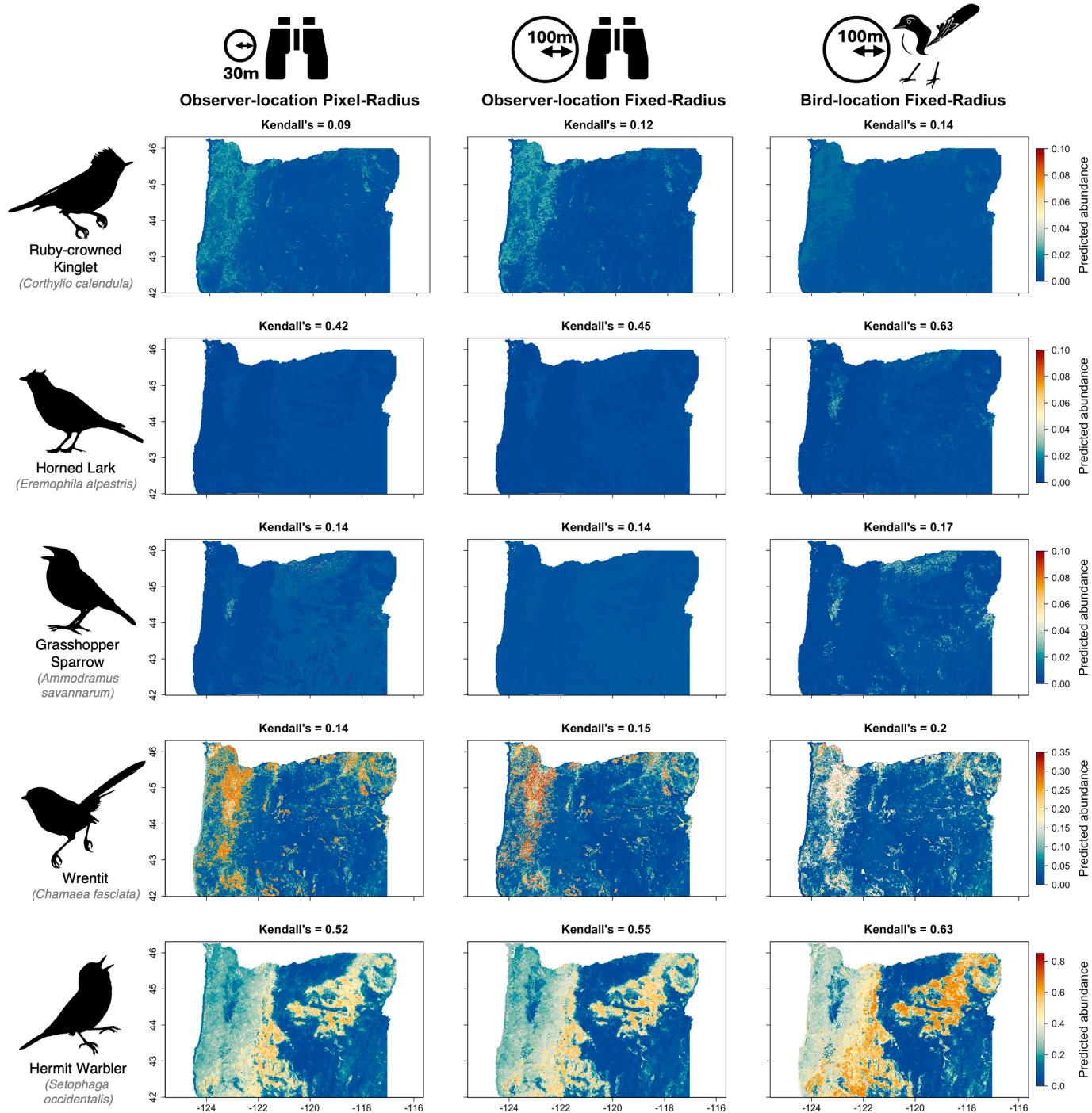


Fig. 5. Predicted relative abundance of five habitat specialists' Poisson BRT models from three habitat characterization strategies in the state of Oregon, USA. Trained models were predicted in the 2019 Landsat raw spectral bands composite. Kendall's rank correlation coefficient is presented on top of each panel. Higher Kendall's values indicate better model performance. The observer-location models refer to habitat characterization performed on observer's location; the bird-location models refer to habitat characterization performed on each bird's estimated location from the distance and direction data obtained during point counts. Fixed radius is a 100 m buffer around location, while pixel radius is 30 m around location. The unit for predicted relative abundance is on a 30×30 m pixel scale. Note: predictions extend across Oregon, even outside each species' known ranges, simply to illustrate the relative differences in the model predictions.

inaccurate data can still be ecologically interpretable. However, small errors can affect interpretation when propagated across large regions, such as in our maps of predicted relative abundances across the state of Oregon. Furthermore, data for habitat specialists, where niche breadth is much narrower than for generalists, may be more sensitive to positional error because habitat differences between observer- and bird-locations may be more likely to differ. In our study, we used Landsat imagery with a fine 30 m spatial resolution to characterize habitats. As

mentioned in Moudrý et al. (2024), when spatial resolution of environmental layers is finer than positional uncertainty, SDMs can result in misleading species-environmental relationships. Thus, our results further support the idea that consideration of positional errors is important to reduce chances of creating misleading species distribution prediction maps. In addition, our results showed that habitat specialists had a higher model performance when using bird-location models than did generalists. These results align with Gábor et al. (2020) who

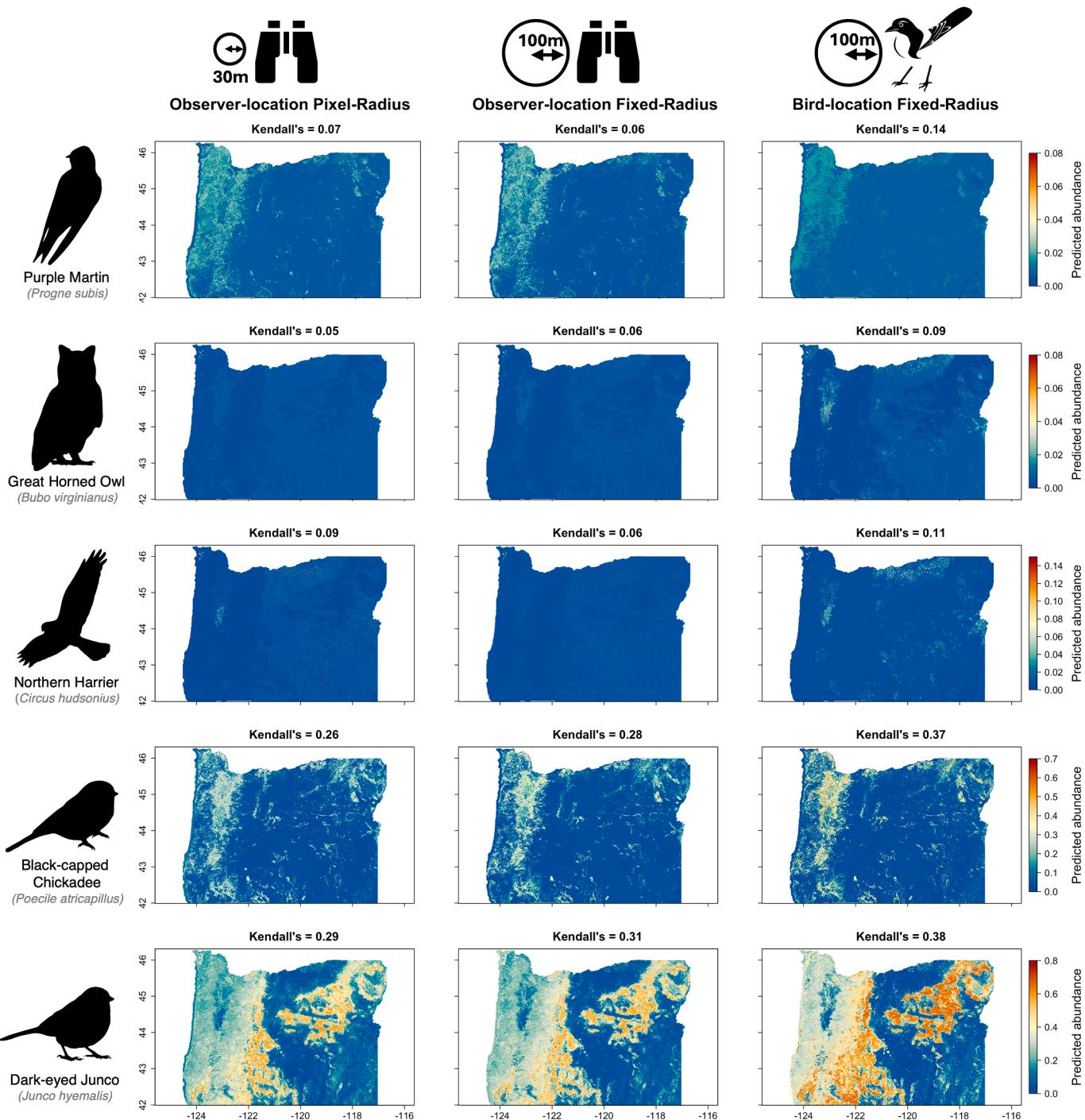


Fig. 6. Predicted relative abundance of five habitat generalists' Poisson BRT models from three habitat characterization strategies in the state of Oregon, USA. Trained models were predicted in the 2019 Landsat raw spectral bands composite. Kendall's rank correlation coefficient is presented on top of each panel. Higher Kendall's values indicate better model performance. The observer-location models refer to habitat characterization performed on observer's location; the bird-location models refer to habitat characterization performed on each bird's estimated location from the distance and direction data obtained during point counts. Fixed radius is a 100 m buffer around location, while pixel radius is 30 m around location. The unit for predicted relative abundance is on a 30×30 m pixel scale. Note: predictions extend across Oregon, even outside each species' known ranges, simply to illustrate the relative differences in the model predictions.

evaluated the effect of introducing positional error in both generalists and specialists; their results showed specialists were more negatively affected by positional error than generalists.

The median detected distance of individual birds across all our distance data was 80 m (Fig. S2). Even when employing a relatively large fixed radius of 100 m for our comparisons of observer- versus bird-location habitats, our models showed better performance when centered on bird locations. That is, for the fixed-radius method, the

radius was large enough, on average, that there should have been a greater degree of overlap when habitats were characterized around each location. Yet, the fixed-radius of 100 m centered on bird-locations outperformed the observer-location fixed-radius models, indicating the Landsat reflectance value from the non-overlapping areas from an observer's locations did differ from that of the bird locations sufficiently to affect results. In many avian surveys, observers typically conduct point count surveys on or near roads. In areas with heterogeneous habitat

cover, birds may often be detected in habitat that differs from observer's location. Therefore, characterizing habitat based on bird-locations can enhance the model's capability to delineate habitat differences from observer-location, further improving model predictability and accuracy. Conversely, in scenarios where the surveyed area consists of less heterogeneous habitat, there might be minimal differences in Landsat reflectance values between observer- and bird-locations, resulting in very similar model performances. Our results from comparisons of generalist species' maps suggests this is the case.

Although our goal was to evaluate model performance between bird- and observer-locations, we also compared the effects of different radii sizes on model accuracy. Generally, fixed-radius models performed better than pixel-radius (30 m) and species-specific effective radius models. Since we only tested three different radii, and species respond differently across gradient sizes of scales (Moudry et al., 2023; Connor et al., 2018), we may not provide comprehensive recommendations on the most reliable radius for characterizing habitat. Here, we briefly speculate on possibilities as to why pixel-radius and effective-radius performed less well than fixed-radius models. For pixel-radius models, precise locations are rarely known because most birds are heard, not seen. Potential errors in the distance and directional estimates may introduce noise in the characterization of environmental predictors. For example, birds are less detectable as the distance from the observer increases (Edwards et al., 2023; Anderson et al., 2015), and this includes uncertainty of location (i.e., positional error) (Moudry et al., 2024). In addition to positional errors resulting from distance estimates, several factors have been identified to impact position estimates. These include biases in global navigation satellite systems, mismatched coordinate reference systems, and rounding of coordinates (Moudry et al., 2024). With a fine 30-m spatial resolution to summarize predictors, mismatching the bird's exact location from a pixel-based raster may result in higher model stochasticity, contributing to lower model performance, since adjacent pixels may have different Landsat reflectance values. For effective radius models, the larger the effective sampling radius, the more overlap exists in comparisons of observer- and bird-location habitats. Therefore, for many species effectively detected at large distances, shifting analyses to bird locations and characterizing habitats based on effective radii should have minimal benefits to model performance. These results are consistent with that of Gábor et al. (2022) who found that coarsening spatial resolution did not fully compensate for the effects of positional errors. Regarding effective radius estimates, we estimated species-specific effective radii from a substantial dataset of distance values collected in Oregon. As with all surveys of birds, distance estimation errors are always present (and the magnitudes of those errors are typically unknown). We recommend that careful evaluation of species-specific effective radii is necessary before use of those radii as buffers to characterize habitats because of occasionally unusual effects of species detectability on effective detection radii.

Many studies have effectively utilized fine-scale (30 m) remotely sensed data combined with larger radii (>100 m) sampled at survey locations to analyze how species respond to complex landscapes (Gábor et al., 2024; Moraga et al., 2019). Despite these larger radii encompassing most habitats that species may occupy, we argue that using such extensive distances (300–2000 m) to summarize habitat characteristics will naturally include habitats that birds do not actually occupy, particularly in more heterogeneous matrix landscapes. Connor et al. (2018) conducted an investigation using seven different grain sizes in SDMs for two virtual species and discovered that the area under the receiver operating characteristic curve (AUC) values generally decreased as grain size increased. Cohen and Jetz (2025) fitted SDMs for 572 bird species across five spatial grains from 1 to 50 km and found that 1 km spatial grain yielded more model accuracy. In contrast, coarsening spatial resolution, however, does not always substantially impact model performance (Pradervand et al., 2014; Guisan et al., 2007). A similar concept used in SDMs – the species-centered approach, often employs a large radius (2 km) to characterize habitat and has been effectively used

to quantify the influence of suitable habitat on bird occupancy (Betts et al., 2014). Nevertheless, applying the species-centered approach to predict potential distributions in a new landscape presents challenges if species' relevant habitats are not specifically characterized. Our study supports the concept that habitat characterization centered more exactly on a detected individual's instead of a surveyor's location assists in capturing the species-habitat relationships and will permit better prediction of distributions across landscapes with similar habitats present. It is common among bird survey protocols, in particular, to gather distance and direction information so the application of mapping bird locations first then collecting habitat data should improve bird SDMs if our results hold in other regions. Given the high habitat diversity of our study region, Oregon, and the large number of species we studied ($n = 105$), it is likely our results are generalizable at least to other North American landscapes.

In the case of habitat specialist and generalist comparisons, a higher percentage of specialists showed better model percentage improvement in bird-location models than generalists. This result is consistent with other studies (Connor et al., 2018; Goedecke et al., 2020; Morelli et al., 2024). Habitat specialists tend to occupy specific, restricted environments with narrower niche breadths, which allows modeling algorithms to more easily delineate differences across diverse landscapes, thereby enhancing the accuracy of the models. This has been validated through both empirical studies (Morelli et al., 2024; Barnagaud et al., 2012) and virtual species simulations (Pérez-Granados and Traba, 2021). Our results further suggest that the occurrences of habitat specialists are strongly correlated with the environmental predictors surrounding each bird's estimated location. It is important to note that specialists with large home ranges, such as the American Goshawk (*Astur atricapillus*), may benefit less from shifting habitat characterization from observer to bird locations, because an individual may occupy a large area. As a result, using a relatively smaller area, such as a 100-m radius, to characterize surrounding predictors may be less effective for these large home range species. Consequently, the arguments presented in this study are more applicable to species with smaller home ranges. In contrast, habitat generalists typically occupy a broad range of habitat gradients (i.e., heterogeneous mosaics), making it more challenging for modeling algorithms to accurately capture the species' associated habitats (Grenouillet et al., 2011).

Our incorporation of relative abundance into models was intended, in part, to illustrate the consequences of small differences in model performance by predicting abundances across the large state of Oregon, even to areas where the birds are not known to breed. This approach applies what is learned about habitats where birds occur and where they were not detected, then ignores details of biogeographic distributional limits to illustrate in a heuristic manner the consequences of model choices. Our approach revealed some interesting differences in predicted distributions and abundances. On average, the predicted relative abundances of habitat specialists were qualitatively similar across our modeling approaches. An exception was Wrentit (*Chamaea fasciata*), where we found obvious discrepancies between predictions derived from bird-location models and those from observer-location models. The observer-location models predicted high relative abundance in metropolitan areas, such as Portland, Oregon, while the bird-location fixed-radius models indicated relatively lower abundance. Wrentits are known to inhabit coastal and interior shrub habitats, but are rarely found in urban settings. As a result, observer-location based model predictions may mislead stakeholders in map interpretation, thereby limiting our ability to inform effective conservation planning. Habitat specialists are often of the greatest conservation concern, requiring more focused efforts for habitat and population restoration (Chazdon et al., 2011). Therefore, our findings highlight that more accurate model predictions can be achieved through bird-location models that accommodate positional error for habitat specialists.

Predicting species' distributions in future habitat or climate scenarios are current foci of ecological research and conservationists are

interested in applying results to management decisions. Nevertheless, caution is warranted because SDMs have limitations when interpolating or extrapolating beyond locations where species occur, as the modeling procedure is based on habitat characterizations that have limits regarding the specificity with which they can currently identify habitat components. For example, our prediction maps of Hermit Warbler (*Setophaga occidentalis*) reveal model insensitivity to different assemblages of tree species forming the coniferous habitats Hermit Warblers choose for breeding. We incorporated raw Landsat spectral bands that depend solely on surface reflectance (numerical data) rather than land cover type (categorical data) when fitting our SDMs. The raw spectral bands yield reflectance values ranging from 0 to 255 for each designated band (Landsat 7 Data Users Handbook). Tree species sharing similar reflectance values across these bands, or those tree species clustered as mixed forests within 30-m pixel, may be less distinguishable using Landsat raw spectral bands. We speculate that the reflectance values obtained from dominant coniferous trees in the western Cascades – Douglas-fir (*Pseudotsuga menziesii*) may closely resemble those of eastern Cascades coniferous species – Ponderosa Pine (*Pinus ponderosa*), Lodgepole Pine (*Pinus contorta*), and Western White Pine (*Pinus monticola*) (see Fig. S9 for all six Landsat spectral bands). Hermit Warblers inhabiting coniferous forests in the western Cascades are in areas where the Landsat-derived reflectance values yield similar results for coniferous forests in the eastern Cascades, leading to predictions of relatively high abundance in eastern Oregon. Further research is needed to investigate the effectiveness of using raw spectral bands from Landsat to accurately characterize different assemblages of tree species for SDMs. Despite the advantages of fitting models on detected individuals instead of observer locations, these models were unable to fully capture the realized niches of the species. Although we utilized the complete range of Landsat spectral bands (i.e., blue, green, and red) to characterize habitat, relying solely on these spectral bands may not adequately differentiate landscape composition, including variations in tree species. In 2030, the anticipated launch of Landsat Next is expected to provide finer spatial resolutions (10–20 m) and temporal scales (6-day interval), along with the addition of 15 new spectral bands (Next, 2023). By integrating land cover type with these advancements, scientists should be better at assessing species-habitat relationships at more refined spatial and temporal scales.

Our analyses included a diverse set of avian species (n = 105), with each species distribution model being trained using six BRT models that responded to various environmental gradients through remote sensing imagery across Oregon. We also compared the performance of these models based on contrasting our categorization of species as habitat generalists versus specialists. However, our study does not encompass the full natural distribution of species (e.g., North America), nor does it incorporate the full range of climatic or landcover type predictors. Thus, we limit our conclusion to the spatial extent, predictors, and species we studied. Our findings demonstrated that the implementation of bird-location fixed-radius models yielded higher accuracy in model predictions. In Europe, the “Ornitho” database (e.g., <https://www.ornitho.ch>) enables observers to submit checklists that include the exact locations of birds. This is a valuable resource as it can assist in bird-location SDMs fitting, especially for habitat specialists. We encourage current avian population monitoring programs to incorporate features that enable observers to submit checklists along with the location of birds. Accounting for positional error in birds can be challenging, particularly for rare or inconspicuous species with low detectability. Autonomous recording units (ARUs) have improved the detectability of birds, enabling scientists to estimate density of population more accurately. Recent studies indicate that in some circumstances ARUs are effective tools for determining bird locations (Mennill et al., 2006; Manzano-Rubio et al., 2022; Frommolt and Tauchert, 2014).

5. Conclusion

This study highlighted differences in SDMs’ performance when characterizing habitat using locations of surveyed individuals compared with observers’ locations. Our research on birds demonstrates that characterizing habitat around bird locations with a 100 m radius generally enhanced model accuracy, particularly for habitat specialists, compared to characterizing habitats around observer locations. The predictions generated by bird-location models provide a more reliable estimate of relative abundance and distribution but must also incorporate biogeographical range information to limit predictions to the known ranges of birds prior to use by conservationists or managers. Our findings emphasize the importance of prioritizing habitat characterization that focuses directly on surveyed individuals’ locations instead of on observer locations. Thus, if direction and distance are available in a dataset, we encourage researchers to prioritize characterizing predictors on location of surveyed individuals. Despite this study mainly focusing on birds, we speculate that our approach can apply to other taxa as well. By implementing these more accurate modeling strategies, we can enhance our understanding of species distribution and abundance.

CRediT authorship contribution statement

Fang-Yu Shen: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Conceptualization. **Fiona Victoria Stanley Jothiraj:** Writing – review & editing, Validation, Software, Methodology, Formal analysis. **Rebecca A. Hutchinson:** Writing – review & editing, Methodology, Conceptualization. **Tyler A. Hallman:** Writing – review & editing, Investigation, Conceptualization. **Jenna R. Curtis:** Writing – review & editing, Investigation. **W. Douglas Robinson:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Fang-Yu Shen reports financial support was provided by Google LLC. Rebecca A. Hutchinson reports financial support was provided by National Science Foundation. W. Douglas Robinson reports financial support was provided by Bob and Phyllis Mace Professorship. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was supported by the Bob and Phyllis Mace Professorship and Hatch Funds (WDR), the Google Earth Engine Uplift grant (FYS), and in part by the National Science Foundation (NSF) [grant number III-2046678] (RAH). A special thank goes to Robert Yelle for assisting with analyses on high performance computing. We are grateful to Ivan Arismendi and two anonymous reviewers for comments that improved this manuscript. We thank all participants who contributed checklists to the Oregon 2020 Birds project.

6. Data accessibility statement

Code used in this study is available online at https://github.com/BettyShen/SDM_bird_observer. The Oregon 2020 data is available in the online supplementary file.

7. Declaration of generative AI in scientific writing

During the preparation of this work, the authors used Grammarly to

improve the readability and language of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ecolind.2025.113546>.

Data availability

Data and code are available in the data accessibility statement

References

- Adde, A., et al., 2023. N-SDM: a high-performance computing pipeline for Nested Species Distribution Modelling. *Ecography* 2023, e06540.
- Amundson, C.L., Royle, J.A., Handel, C.M., 2014. A hierarchical model combining distance sampling and time removal to estimate detection probability during avian point counts. *Auk Ornithol. Adv.* 131, 476–494.
- Anderson, A.S., Marques, T.A., Shoo, L.P., Williams, S.E., 2015. Detectability in Audio-Visual Surveys of Tropical Rainforest Birds: The Influence of Species, Weather and Habitat Characteristics. *PLOS ONE* 10, e0128464.
- Araújo, M.B., et al., 2019. Standards for distribution models in biodiversity assessments. *Sci. Adv.* 5, eaat4858.
- Bagchi, R., et al., 2013. Evaluating the effectiveness of conservation site networks under climate change: accounting for uncertainty. *Glob. Change Biol.* 19, 1236–1248.
- Barbet-Massin, M., Jetz, W., 2014. A 40-year, continent-wide, multispecies assessment of relevant climate predictors for species distribution modelling. *Divers. Distrib.* 20, 1285–1295.
- Barnagaud, J.-Y., et al., 2012. Relating Habitat and Climatic Niches in Birds. *PLoS ONE* 7, e32819.
- Betts, M.G., et al., 2014. A species-centered approach for uncovering generalities in organism responses to habitat loss and fragmentation. *Ecography* 37, 517–527.
- Betts, M.G., et al., 2022. Forest degradation drives widespread avian habitat and population declines. *Nat. Ecol. Evol.* 6, 709–719.
- Buckland, S.T., Anderson, D.R., Burnham, K.P., 2007. *Advanced Distance Sampling*. Oxford University Press.
- Buckland, S.T., Rexstad, E.A., Marques, T.A., Oedekoven, C.S., 2015. *Distance Sampling: Methods and Applications*. Springer International Publishing, Cham. <https://doi.org/10.1007/978-3-319-19219-2>.
- Burns, P., et al., 2020. Incorporating canopy structure from simulated GEDI lidar into bird species distribution models. *Environ. Res. Lett.* 15, 095002.
- Chazdon, R.L., et al., 2011. A novel statistical method for classifying habitat generalists and specialists. *Ecology* 92, 1332–1343.
- Clements, N.M., Shen, F.-Y., Robinson, W.D., 2025. A 50-Year Perspective on Changes in a Pacific Northwest Breeding Forest Bird Community Reveals General Stability of Abundances. *Diversity* 17, 123.
- Cohen, J.M., Jetz, W., 2025. Fine-Grain Predictions Are Key to Accurately Represent Continental-Scale Biodiversity Patterns. *Glob. Ecol. Biogeogr.* 34, e13934.
- Connor, T., et al., 2018. Effects of grain size and niche breadth on species distribution modeling. *Ecography* 41, 1270–1282.
- Crego, R.D., Stabach, J.A., Connette, G., 2022. Implementation of species distribution models in Google Earth Engine. *Divers. Distrib.* 28, 904–916.
- Diefenbach, D.R., Brauning, D.W., Mattice, J.A., 2003. Variability in Grassland Bird Counts Related to Observer Differences and Species Detection Rates. *The Auk* 120, 1168–1179.
- Edwards, B.P.M., et al., 2023. Point count offsets for estimating population sizes of north American landbirds. *Ibis* 165, 482–503.
- Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. *J. Anim. Ecol.* 77, 802–813.
- Farmer, R.G., Leonard, M.L., Horn, A.G., 2012. Observer effects and avian-call-count survey quality: rare-species biases and overconfidence. *The Auk* 129, 76–86.
- Farnsworth, G.L., et al., 2002. A removal model for estimating detection probabilities from point-count surveys. *The Auk* 119, 414–425.
- Fernandez, M., et al., 2009. Locality uncertainty and the differential performance of four common niche-based modeling techniques. *Biodivers. Inform.* 6, 36–52.
- Frommolt, K.-H., Tauchert, K.-H., 2014. Applying bioacoustic methods for long-term monitoring of a nocturnal wetland bird. *Ecol. Inform.* 21, 4–12.
- Gábor, L., et al., 2020. The effect of positional error on fine scale species distribution models increases for specialist species. *Ecography* 43, 256–269.
- Gábor, L., et al., 2022. Positional errors in species distribution modelling are not overcome by the coarser grains of analysis. *Methods Ecol. Evol.* 13, 2289–2302.
- Gábor, L., et al., 2023. Species distribution models affected by positional uncertainty in species occurrences can still be ecologically interpretable. *Ecography* 2023, e06358.
- Gábor, L., Cohen, J., Moudré, V., Jetz, W., 2024. Assessing the applicability of binary land-cover variables to species distribution models across multiple grains. *Landscape Ecol.* 39, 66.
- Gaston, K.J., Blackburn, T.M., Goldewijk, K.K., 2003. Habitat conversion and global avian biodiversity loss. *Proc. r. Soc. Lond. B Biol. Sci.* 270, 1293–1300.
- Goedecke, F., Marcenò, C., Guarino, R., Jahn, R., Bergmeier, E., 2020. Reciprocal extrapolation of species distribution models between two islands – Specialists perform better than generalists and geological data reduces prediction accuracy. *Ecol. Indic.* 108, 105652.
- Goetz, S.J., Sun, M., Zolkos, S., Hansen, A., Dubayah, R., 2014. The relative importance of climate and vegetation properties on patterns of North American breeding bird species richness. *Environ. Res. Lett.* 9, 034013.
- Grenouillet, G., Buisson, L., Casajus, N., Lek, S., 2011. Ensemble modelling of species distribution: the effects of geographical and environmental ranges. *Ecography* 34, 9–17.
- Guisan, A., Graham, C.H., Elith, J., Huettemann, F., Group, the N. S. D. M., 2007. Sensitivity of predictive species distribution models to change in grain size. *Divers. Distrib.* 13, 332–340.
- Guisan, A., Thuiller, W., 2005. Predicting species distribution: offering more than simple habitat models. *Ecol. Lett.* 8, 993–1009.
- Hallman, T.A., Robinson, W.D., 2020. Comparing multi- and single-scale species distribution and abundance models built with the boosted regression tree algorithm. *Landscape Ecol.* 35, 1161–1174.
- Hallman, T.A., Robinson, W.D., 2020. Deciphering ecology from statistical artefacts: Competing influence of sample size, prevalence and habitat specialization on species distribution models and how small evaluation datasets can inflate metrics of performance. *Divers. Distrib.* 26, 315–328.
- Hopkins, L.M., Hallman, T.A., Kilbride, J., Robinson, W.D., Hutchinson, R.A., 2022. A comparison of remotely sensed environmental predictors for avian distributions. *Landscape Ecol.* 37, 997–1016.
- Jan, A., Arismendi, I., Giannico, G., 2025. Double Trouble for Native Species Under Climate Change: Habitat Loss and Increased Environmental Overlap With Non-Native Species. *Glob. Change Biol.* 31, e70040.
- Jefferys, K., et al., 2024. Breeding habitat loss linked to declines in Rufous Hummingbirds. *Avian Conserv. Ecol.* 19, art.2.
- Landsat 7 Data Users Handbook | U.S. Geological Survey. <https://www.usgs.gov/media/files/landsat-7-data-users-handbook>.
- Manzano-Rubio, R., Bota, G., Brotons, L., Soto-Largo, E., Pérez-Granados, C., 2022. Low-cost open-source recorders and ready-to-use machine learning approaches provide effective monitoring of threatened species. *Ecol. Inform.* 72, 101910.
- Marsh, H., Sinclair, D.F., 1989. Correcting for visibility bias in strip transect aerial surveys of aquatic fauna. *J. Wildl. Manag.* 1017–1024.
- Marshall, D.B., Hunter, M.G., Contreras, A., 2003. *Birds of Oregon: A General Reference*. Oregon State University Press, Corvallis Or.
- Matuoka, M.A., Benchimol, M., de Almeida-Rocha, J.M., Morante-Filho, J.C., 2020. Effects of anthropogenic disturbances on bird functional diversity: a global meta-analysis. *Ecol. Indic.* 116, 106471.
- Mennill, D.J., Burt, J.M., Fristrup, K.M., Vehrencamp, S.L., 2006. Accuracy of an acoustic location system for monitoring the position of duetting songbirds in tropical forest. *J. Acoust. Soc. Am.* 119, 2832–2839.
- Mitchell, M.S., Lancia, R.A., Gerwin, J.A., 2001. Using Landscape-Level Data to Predict the Distribution of Birds on a Managed Forest: Effects of Scale. *Ecol. Appl.* 11, 1692–1708.
- Moraga, A.D., Martin, A.E., Fahrig, L., 2019. The scale of effect of landscape context varies with the species' response variable measured. *Landscape Ecol.* 34, 703–715.
- Morelli, F., et al., 2024. Which bird traits most affect the goodness-of-fit of species distribution models? *Ecol. Indic.* 158, 111317.
- Moudré, V., et al., 2023. Scale mismatches between predictor and response variables in species distribution modelling: A review of practices for appropriate grain selection. *Prog. Phys. Geogr. Earth Environ.* 47, 467–482.
- Moudré, V., et al., 2024. Optimising occurrence data in species distribution models: sample size, positional uncertainty, and sampling bias matter. *Ecography* 2024, e07294.
- Landsat Next | Landsat Science. <https://landsat.gsfc.nasa.gov/satellites/landsat-next/> (2023).
- Olea, P.P., Mateo-Tomás, P., 2011. Spatially explicit estimation of occupancy, detection probability and survey effort needed to inform conservation planning. *Divers. Distrib.* 17, 714–724.
- Osborne, P.E., Leitão, P.J., 2009. Effects of species and habitat positional errors on the performance and interpretation of species distribution models. *Divers. Distrib.* 15, 671–681.
- Pankratz, R.F., Hache, S., Sólymos, P., Bayne, E.M., 2017. Potential benefits of augmenting road-based breeding bird surveys with autonomous recordings. *Avian Conserv. Ecol.* 12, 18.
- Pérez-Granados, C., Traba, J., 2021. Estimating bird density using passive acoustic monitoring: a review of methods and suggestions for further research. *Ibis* 163, 765–783.
- Pradervand, J.-N., Dubuis, A., Pellissier, L., Guisan, A., Randin, C., 2014. Very high resolution environmental predictors in species distribution models: Moving beyond topography? *Prog. Phys. Geogr.* 38, 79–96.
- Reynolds, R.T., Scott, J.M., Nussbaum, R.A., 1980. A variable circular-plot method for estimating bird numbers. *The Condor* 82, 309–313.
- Robinson, W.D., et al., 2021. Big Bird Plots: Benchmarking Neotropical Bird Communities to Address Questions in Ecology and Conservation in an Era of Rapid Change. *Front. Ecol. Evol.* 9.
- Robinson, W.D., Hallman, T.A., Curtis, J.R., 2020. Benchmarking the avian diversity of Oregon in an era of rapid change. *Northwest. Nat.* 101, 180–193.
- Shen, F.-Y., Ding, T.-S., Tsai, J.-S., 2023. Comparing avian species richness estimates from structured and semi-structured citizen science data. *Sci. Rep.* 13, 1214.
- Shirley, S.M., et al., 2013. Species distribution modelling for the people: unclassified landsat TM imagery predicts bird occurrence at fine resolutions. *Divers. Distrib.* 19, 855–866.

- Smith, A.B., Murphy, S.J., Henderson, D., Erickson, K.D., 2023. Including imprecisely georeferenced specimens improves accuracy of species distribution models and estimates of niche breadth. *Glob. Ecol. Biogeogr.* 32, 342–355.
- Valavi, R., Elith, J., Lahoz-Monfort, J.J., Guillera-Arroita, G., 2019. blockCV: An r package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods Ecol Evol* 225–232. <https://doi.org/10.1111/2041-210X.13107>.
- Vincenty, T., 1975. Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations. *Surv. Rev.* <https://doi.org/10.1179/sre.1975.23.176.88>.
- What are the best Landsat spectral bands for use in my research? | U.S. Geological Survey. <https://www.usgs.gov/faqs/what-are-best-landsat-spectral-bands-use-my-research>.
- Whytock, R.C., et al., 2018. Bird-community responses to habitat creation in a long-term, large-scale natural experiment. *Conserv. Biol.* 32, 345–354.
- Williams, J.J., Newbold, T., 2020. Local climatic changes affect biodiversity responses to land use: A review. *Divers. Distrib.* 26, 76–92.
- Zhang, G., Zhu, A.-X., Huang, Z.-P., Xiao, W., 2018. A heuristic-based approach to mitigating positional errors in patrol data for species distribution modeling. *Trans. GIS* 22, 202–216.
- Ziolkowski Jr., D., Lutmerding, M., Aponte, V. & Hudson, M.-A. 2022 Release - North American Breeding Bird Survey Dataset (1966-2021). U.S. Geological Survey <https://doi.org/10.5066/P97WAZE5> (2022).
- Zurell, D., et al., 2020. A standard protocol for reporting species distribution models. *Ecography* 43, 1261–1277.